Research Paper | Education | E-ISSN No : 2454-9916 | Volume : 6 | Issue : 4 | April 2020

EFFECT OF IMPLEMENTATION OF USER GUIDE ON SELECTION OF APPROPRIATE STATISTICAL TECHNIQUES

Sangita Birajdar

MIT Arts, Commerce and Science College, Alandi, Pune, India 412105.

ABSTRACT

The National Policy on Education 2016, of India in its report stated that, the quality and also quantity of research and innovations in India emerging out of Institutions of higher education and research leaves much to be desired. The statistical illiteracy is the most important reason for bringing down the quality of research in education and social sciences.

The present study is an attempt to enhance the statistical literacy and to develop the statistical thinking among novice researchers in the field of education and social sciences. The main purpose of this research is to investigate whether the use of User Guide for selecting appropriate statistical technique/s improves the ability of novice researchers in data analysis and reduces the common errors committed at the time of selecting appropriate statistical technique/s.

KEYWORDS: User Guide, Statistical techniques, descriptive statistics, inferential statistics, Control Group, Experimental Group.

INTRODUCTION:

"Statistics is a branch of scientific methodology. It deals with the collection, classification, description and interpretation of data obtained by conducting surveys and experiments. Its essential purpose is to describe and draw inferences about numerical properties of populations" (Ferguson, 1966).

The wide use of subject Statistics in almost all fields while doing research needs knowledge about statistical techniques. Different statistical techniques have been preferred in the research of different disciplines. Educational research is the systematic application of scientific methods for solving the educational problems. It uses both conceptions of social reality and the statistical methods that are considered appropriate for exploring it.

Statistical methods are often used to communicate research findings and to support hypotheses and give reliability to research methodology and conclusions in educational researches. Statistical methods are the main pillars for solutions of many issues in educational process in which the teachers, students, parents and administrators are interested. Incorrect choice of a method of the experimental data analysis can lead to erroneous conclusions, incorrect interpretation of the research results, and thereby distort or even lead to the loss of the scientific value of such research results and the loss of informativity (Khusainova et al. 2016). Statistical methods are important aids to detect trends, explore relationships and draw conclusions from experimental data. However, it is not uncommon to find that many researchers apply statistical tests without first checking whether they are appropriate for the intended application (Granato 2014).

Selection of an appropriate statistical method for the data under study is an important task in the entire research process. For acquiring the mastering skills of competent knowledge application in various research activities needs proper guidance. For instance, suppose the data is collected to find the effect of safety program given for Chemistry laboratory, the paired t-test will be suitable for this if the collected sample before and after safety program get implemented comes from normal population.

There are plenty of statistical software packages like SAS, MINITAB, MATLAB, SPSS, etc. available for data analysis but they are less convenient and costly. In the present study MS-EXCEL is selected for data analysis as all of us have easy access to Excel on our own computers and do not need to invest in other software. Excel is not only useful for data presentation by means of tables, diagrams and graphs but also it is useful for descriptive and inferential research. As a spread sheet, Excel can be used for data entry, manipulation and presentation but it also offers a set of statistical analysis functions and other tools that can be used to run descriptive statistics and to perform several different and useful inferential statistical tests that are widely used in educational and social sciences researches.

Taking into consideration the importance of statistics in educational research, a study is conducted in which the User Guide is prepared for guiding novice researchers on choosing appropriate statistical technique/s for data analysis using Excel. It guides on choosing appropriate statistical techniques required in the researches of social sciences and education in an easy and understandable way. It facilitates the researchers for giving interpretation on their data.

The present study aims to investigate the potential of User Guide for selecting appropriate statistical techniques for data analysis on the basis of scores obtained in post-test, time required to complete the task and percentage of common errors committed by the novice researchers while analysing the data.

LITERATURE REVIEW:

(Balson, 1959) proposed the statistical techniques for educational research. He emphasized the need of using appropriate statistical technique for educational research by giving the example of the experimental study of education in Australia. The experimental study of education in Australia has been hampered by the failure of teacher-training institutions to provide for the presentation of suitable statistical techniques which the average student is capable of acquiring and by the type of researcher whose published studies in the field of education are of doubtful value because basic assumptions underlying the statistical techniques used in the investigations have not been considered or met.

(Skidmore & Thompson, 2010) provided a historical account and Meta synthesis of which statistical techniques are most frequently used in the fields of education and psychology. They reviewed six articles from the American Educational Research Journal from 1969 to 1997 and five articles from the psychological literature from 1948 to 2001 resulted in a total number of 17,698 techniques recorded from the 12,012 articles reviewed. They discussed the trends for the education and psychology literature both individually and collectively.

Statistical errors are common in scientific literature, and about 50% of the published articles have at least one error. Many of the statistical procedures including correlation, regression, t- tests, and analysis of variance, namely parametric tests, are based on the assumption that the data follows a normal or a Gaussian distribution. Normality and other assumptions should be taken seriously, for when these assumptions do not hold, it is impossible to draw accurate and reliable conclusions about reality (Ghasemi and Zahediasl, 2012).

Each and every researcher should have some knowledge in Statistics and must use statistical tools in his or her research. One should know about the importance of statistical tools and how to use them in their research or survey (Begum and Ahmed, 2015). They discussed the importance of statistical tools (Here the tools discussed are nothing but techniques/ methods of data analysis) and gave a report on statistical tools used in research studies. According to them, "Simple inspection of data, without statistical treatment, by an experienced and dedicated analyst may be just as useful as statistical figures on the desk of the disinterested."

The quality assurance of the research work must be dealt with the appropriate statistical operations as making mistakes in analytical work is unavoidable. (Gupta et al., 2015) carried out study for finding the suitability of statistical methods used in the analysis of the data in Ph.D. theses of social science faculty in Indian Universities. Their analysis reveals a serious and pathetic situation of the status of research in the country.

(Arkkelin, 2014) in his book provided an introduction to how to use the Statistical Package for the Social Sciences (SPSS) for data analysis. The text includes step-by-step instructions, along with screen shots and videos, to conduct various procedures in SPSS to perform statistical data analysis. (Alvi, 2016) prepared a manual for selecting sampling techniques in research which describes what techniques are most suitable for the various sorts of researches.

Begum & Ahmed (2015) suggested the use of excel as an important statistical tool for data analysis. They also suggested the use of data analysis tool for Random Number Generation, to test a hypothesis in Excel to analyse data. Excel's data analysis capabilities make it possible to conduct some advanced analyses of survey data but not others However a program known as XL Stat expands the analytical capabilities of Excel. Tools such as SAS and SPSS are designed with research professionals in mind and make a full range of analytical methods possible.

An exclusive literature survey related to most frequently used statistical techniques by educational and social sciences researchers (Balson, 1959; Karada, 2010; Skidmore & Thompson, 2010) were carried out. In their research they analysed research articles, masters' work, and doctoral work to take an account about the importance of statistical techniques and most frequently used statistical techniques by the researchers in their researches. (Khusainova et al., 2016 provided an algorithm that allows choosing a valid method of statistical data processing. While (Gupta et al, 2015) found the suitability of statistical methods used in the analysis of the data in Ph.D. theses of social science and education faculty in Indian Universities. They also found the types of common errors committed by the researchers while using statistical methods to analyze the data in their Ph.D. theses. (Keselman 1998; Begum & Ahmed, 2015) suggested the use of statistical software package for data analysis. The Handbooks / Manuals were prepared (Arkkelin, 2014; Alvi, 2016) for different purposes.

METHODOLOGICAL FRAMEWORK AND RESEARCH DESIGN:

In order to investigate the effect of implementation of User Guide for selecting appropriate statistical technique/s for data analysis, the User Guide manual was developed by the researcher. It contained the information on using statistical techniques available in the Microsoft Excel for data analysis. It also guided on choosing appropriate statistical techniques required in the researches of social sciences in an easy and understandable way. The User Guide was validated by a team of three statisticians and the suggestions given by this team were incorporated in the User Guide manual.

A purposive sample of size 50 was selected from the target population which was randomly divided into two groups (Control and Experimental groups). Quasi-Experimental design was used in this study. The sample included novice researchers from social sciences and education who were teachers from MIT junior college (16%) and teachers from MIT senior college (64%), student-teachers (who are teaching in schools) and doing their B.Ed. from MIT Saint Dnyneshwar B.Ed. College, Alandi (30%).

Among these 50 novice researchers all have the basic knowledge of using MS-Excel; only 5 of them had a little knowledge of using MS-Excel for data analysis. 22 participants had learned Statistics as one of the subjects at graduation level. So these participants had the basic knowledge about statistical techniques and these were selected in Control Group but 28 participants were either from Arts background or even if they did have a science background they had not taken statistics as one of the subjects. In this study it was important to develop the ability of data analysis among these 28 participants. Hence, these participants were selected in Experimental Group.

Both the groups were taught data analysis techniques by using MS-Excel during the same time by demo method. A post- test was conducted after two weeks to check the knowledge of the researchers regarding the use of MS-Excel for data analysis and choosing an appropriate statistical technique. The Control Group was assessed for the use of MS-Excel for data analysis and choosing appropriate statistical technique/s without using User Guide. However, for the Experimental Group the same task was given and assessed in exactly the same way as the control group, but this group was permitted to use User Guide as a learning aid. The quantitative research approach was emphasized objectivity in measuring the effectiveness of User Guide as a learning aid.

The diagnostic test was prepared by the researcher and validated by the experts to check if the test adhered to the assessment guidelines and also assess its level of difficulty in accordance to the assessment guidelines. The diagnostic test was marked according to the level of difficulty. A full, partial, or no mark was allocated according to the specifications of the memorandum. Each participant had assigned three data sets and they had to decide appropriate statistical technique/s for analyzing data and to carry out the analysis by choosing correct data analysis tool from Ms-Excel and finally to make the correct interpretation. Total 80 scores were assigned for entire task and for every step scores were pre decided. Five excel sheets were assigned randomly to all participants from both groups. The test was administered to all participants. No time limit was given to complete the task. As discussed earlier the participants from Experimental Group were provided the User Guide as learning aid for completing their task whereas the participants from Control Group have to complete their task on the basis of the knowledge they acquired during demo lecture. The time required for completion of test was recorded and the performance of each participant was assessed on the basis of scores pre assigned.

While selecting statistical techniques for data analysis the participants from both the groups committed some common errors like selection of simple bar diagram when pie diagram was appropriate, used t-test without testing equality of population variances, preformed regression analysis without checking the relation between given variables is linear or not etc.. The percentage of common errors committed by the participants while choosing appropriate statistical technique/s was recorded.

RESEARCH QUESTIONS AND HYPOTHESES:

1) Do the educational and social sciences researchers use the appropriate statistical methods/techniques for the analysis of data in their research?

2) What are the types of common errors committed by the researchers while using statistical methods for data analysis?

3) Whether the User Guide is useful for the novice researchers for data analysis in their research work?

H_{01}: There is no significant difference in the average scores of control and experimental groups on selecting appropriate statistical techniques.

H_{02}: There is no significant difference in the average time required for completing the task for control group and experimental group.

ANALYSIS OF SCORES OBTAINED IN POST-TEST:

The quantitative data generated in the post-test for answering the research questions and hypotheses stated above was analysed by using descriptive statistics methods and inferential methods. The use of descriptive statistics is the most fundamental way to summarize data. Descriptive statistics (sometimes referred as summary statistics) are thus used to summarize, organize and reduce large numbers of observations (McMillan & Schumacher, 2010) in the data.

The significance of the performances of both the groups with respect to the scores obtained in post-test could be compared by using two sampled t-test. The assumption for the test is that both groups should be sampled from normal distribution with equal variances. The test can be conducted even though the population variances are unequal. But the normality assumption is very important. Hence, the validity of these assumptions was tested here. Anderson-Darling Test is used to test the normality of two samples and after that F-test is used to test the equality of two population variances.

Descriptive Statistics Method:

Following table summarizes performance of participants from both the groups in terms of scores obtained in post-test, time required (in minutes) to complete the task and percentage of errors committed by both the groups while analysing data using appropriate statistical technique/s. MINITAB 14 package was used to obtain descriptive statistics.

Table 1: Descriptive Statistics on three variables

Descriptive Statistics	Scores obtained in post-test		Time (in minutes) required to complete the task in post-test		% of errors committed by the participants	
	Control Group	Experimental Group	Control Group	Experimental Group	Control Group	Experimental Group
Count	22	28	22	28	22	28
Mean	32.91	57.11	151.55	130.68	46.97	18.45
Standard Error of mean	3.01	1.98	3.33	1.65	4.98	3.25
Standard Deviation	14.12	10.49	15.64	8.71	23.36	17.18
Sample Variance	199.52	110.1	244.64	75.86	545. 92	295.02
Coefficient of variation	42.92	18.37	10.32	6.6	49.74	93.08
Minimum	10	35	120	115	0	0
Q1	20	48.25	139.25	124	33.33	0
Median	33.5	60	149	137.75	50	16.67

Q3	41.5	65	161	137.75	66.67	33.33
Maximum	60	73	180	147	83.33	50
Range	50	38	60	32	83.33	50
Inter Quartile Range	21.5	16.75	21.75	13.75	33.34	33.33
Skewness	-0.04	-0.65	0.18	-0.05	-0.1	0.43
Kurtosis	-0.55	-0.41	-0.37	-0.85	-0.63	-0.99

Figures obtained in Table 1 for both the groups revealed that,

- The average score obtained by the participants of Experimental group is more whereas the average time required for completing the task and average percentage of error committed by the participants of Experimental group is less than that of the participants of Control group.
- The measures of variation indicate that the scores obtained, the time required to complete the task and percentage of error committed by the participants of Control group are more variable than that of Experimental group.
- The Inter Quartile Range indicates that the spread of scores obtained by Control group is more than that of Experimental group.

Analysis of scores by using Inferential Statistics Method:
The hypotheses stated above were tested by using two sampled t-test. The normality assumption for this test was tested by using Anderson-Darling Test and the assumption of equal variances was tested by using F-test.

t-test for testing the hypothesis H_{01}:
The Anderson-Darling Test was conducted by using MINITAB for testing the normality of scores obtained in post-test by the participants of Control Group and Experimental Group. The MINITAB analysis provides a normal probability plot with vertical scale on the graph resembles the vertical scale on normal probability paper. The horizontal axis is a linear scale. The line forms an estimate of the cumulative distribution function for the population from which data are drawn. Numerical estimates of the population parameters, (mean of normal distribution) and (standard deviation of normal distribution), the normality test value of Anderson-Darling test statistic, and the associated p-value are displayed with the plot. The p-value ("probability") is the probability of getting a result that is more extreme if the null hypothesis is true. If the p value is low (e.g., <=0.05), then the data do not follow the normal distribution. The results are displayed in the following figure.

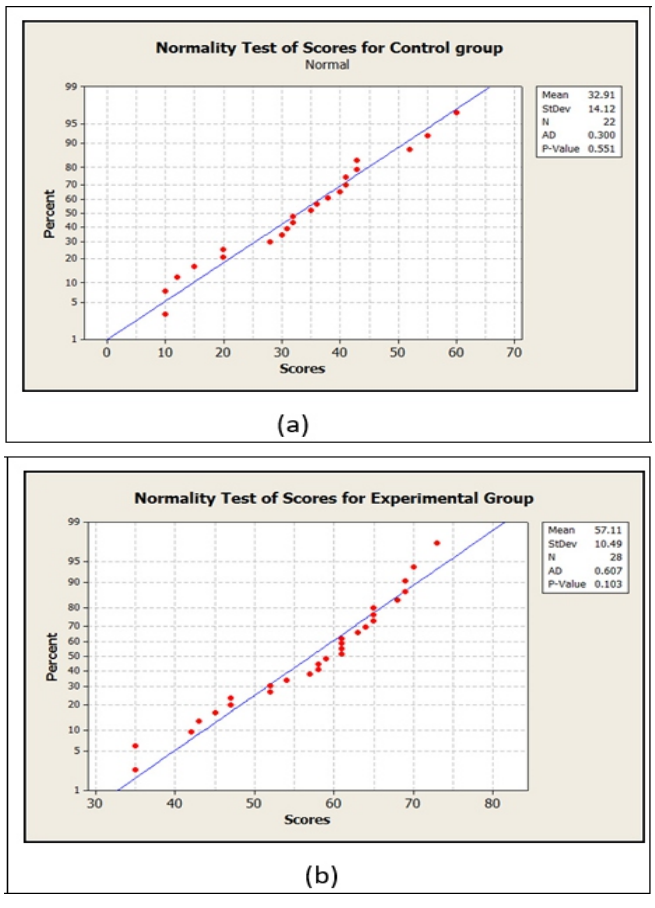

Figure 1: Normality test of scores for Control group and Experimental group

From Figure 1(a) and 1(b), the Anderson-Darling Test reveals that the scores of Control Group as well as Experimental Group are normally distributed.

F-test was conducted to test the equality of variances and it is not significant with p-value 0.146. So the two sampled t-test for testing equality of means with equal population variances was conducted using MINITAB which is highly significant with p-value 0.000 which indicates that the average scores of Experimental Group is more than that of Control Group. Following Box plots show the visual display of average differences of scores and their spreads.

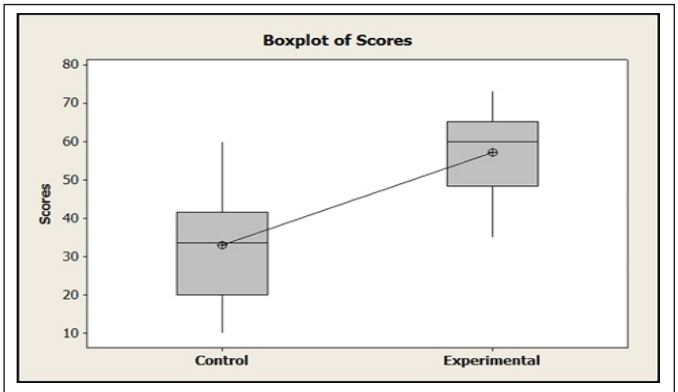

Figure 2: Box plot of Scores for Control and Experimental groups

t-test for testing the hypothesis H_{02}:
Like above, the hypothesis H02 was tested by using t-test. The Anderson-Darling Test was conducted by using MINITAB for testing the normality of time required for completing the task on post-test by the participants of Control Group and Experimental Group. The MINITAB analysis revealed that the time required to complete the task for both the groups is normally distributed as shown in Figure 3(a) and 3(b).

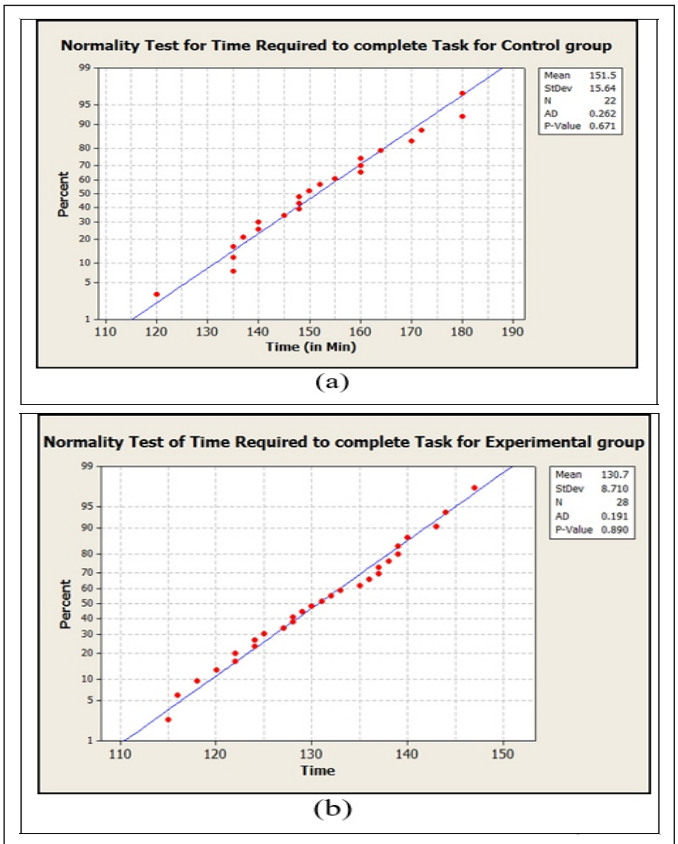

Figure 3: Normality test for time required (in minutes) to complete the task for both groups

F-test was conducted to test the equality of variances and it is significant with p-value 0.005. So the two sampled t-test for testing equality of means with unequal population variances was conducted using MINITAB which is highly significant with p-value 0.000 which indicates that the average time (in minutes) required to complete the task for the participants of Experimental Group is less than that of Control Group. Following Box plots show the visual display of average differences of time required to complete the task and their spreads.

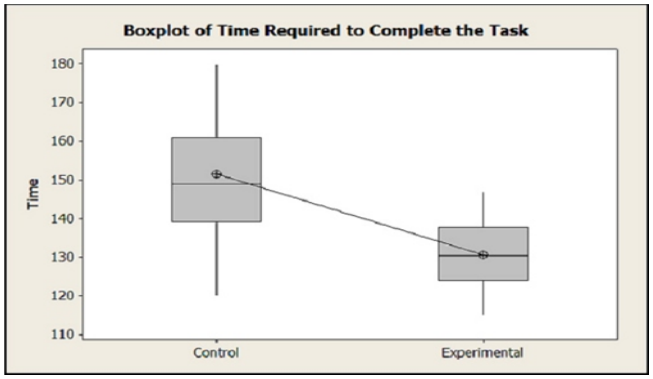

Figure 4: Box plot of time required to complete the task given in post-test for Control and Experimental groups

Analysis of common errors committed by the novice researchers while analyzing data:

While selecting statistical techniques for data analysis the participants from both the groups committed some common errors like selection of simple bar diagram or subdivided bar diagram when pie diagram was appropriate, used t-test without testing equality of population variances, used paired t-test instead of two sampled t-test, preformed regression analysis without checking the relation between given variables is linear or not etc. The analysis of common errors committed by the participants while choosing appropriate statistical technique/s revealed that average percentage of common errors committed by the participants of Control Group is very high as compared to the participants of Experimental Group. Following Box plots show the visual display of average differences of percentage of common error committed by the participants of both the groups and their spreads.

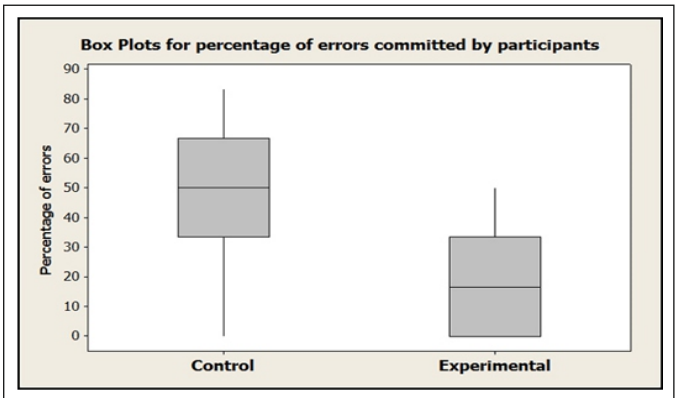

Figure 5: Box Plot of percentage of common error committed by participants while using statistical methods

DISCUSSIONS:

Many people have a fear of calculations and hence they dislike the subjects like Mathematics and Statistics. But interest about the calculation can be generated among these people. Therefore, the software, Ms-Excel which is easily available for everyone was selected in this study. The motive behind this study was when the people learn to get entire calculation on a single click and they will take more interest in statistical analysis. The performance of participants from both the groups was analysed on the basis of quantitative data generated on post-test. This data facilitate examining the research questions asked in earlier section of this paper as follows:

- The participants from experimental group who were allowed to use User Guide for data analysis performed better than the participants from Control group in terms of their scores on the test.
- The participants from Experimental group required less time on an average to complete their task assigned at post-test than the participants from Control group.
- It was observed that the participants from Experimental group on an average committed less percent of common errors while using statistical methods for data analysis than the participants from Control group.

According to (Ghasemi and Zahediasl, 2012), "Statistical errors are common in scientific literature, and about 50% of the published articles have at least one error." (Gupta et al., 2015) carried out study for finding the suitability of statistical methods used in the analysis of the data in Ph.D. theses of social science faculty in Indian Universities. They found that, only 47% of the researchers fulfilled the assumptions of the method used by them. The common errors committed by the researchers while using statistical methods are of serious nature. In the present study the average percentage of common errors committed by the participants of Control group who were not allowed to use User Guide was 46.97%. Hence, this study supports the findings of above two researches. The average percentage of common error committed by the participants of Experimental group was very low 18.45%. This study conclude that, even though the participants have not learned Statistics as one of the subjects at their graduation level, if they use User Guide for their data analysis, they can choose appropriate method and improve the quality of their research.

CONCLUSIONS & RECOMMENDATIONS:

The above results indicate that, the User Guide is useful for the novice researchers from education and social sciences for data analysis in their research work. The use of User Guide saves the time of the researcher and they can interpret, conclude and recommend their findings confidently. It also reduces the common errors generally committed while using statistical methods for data analysis and improves the quality of research. Hence, the use of User Guide is strongly recommended for the novice researchers.

REFERENCES:

1. Alvi M (2016). A Manual for Selecting Sampling Techniques in Research, https://mpra.ub.uni-muenchen.de/70218/
2. Arkkelin D (2014). Using SPSS to Understand Research and Data Analysis, Psychology Curricular Materials. Book, http://scholar.valpo.edu/psych_oer/1
3. Balson M (1959). Statistical Techniques for Educational Research, Australian Journal of Psychology, 11(2), 191-195
4. Begum K. and Ahemed A. (2015), The Importance of Statistical Tools in Research Work, International Journal of Scientific and Innovative Mathematical Research,3(12), 50-58
5. Brancato G., Macchia S. , Murgia M. , Signore M. , Simeoni G. – Italian National Institute of Statistics, ISTAT (2006). Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System, Version 1, https://www.istat.it/it/files/2013/12/Handbook_questionnaire_development_2006
6. Ferguson, G. (1966). Statistical Analysis in Psychology and Education, 2nd Edition, McGraw-Hill, London.
7. Ghasemi and Zahediasl (2012), "Normality Tests for Statistical Analysis: A Guide for Non-Statisticians", International Journal of Endocrinology and Metabolism, 10(2):486-489
8. Granato D., Veronica Calado and Jarvis B.(2014). Observations on the use of statistical methods in Food Science and Technology. Food Research International,55, 137-149.
9. Gupta S et al (2015), Evaluation of Statistical Methods used in Ph.D. Theses of Social Sciences In Indian Universities, International Journal of Recent Scientific Research, 6 (3), 2926 -2931
10. Karada E. (2010). An Analysis of Methods and Statistical Techniques Used By Doctoral Dissertation at the Education Sciences in Turkey. Current Issues in Education, 13(4), 1-20.
11. Keselman, H., Huberty, C. , Lix, L., Olejnik, S., Cribbie, R., Donahue, B., . . . & Levin, J. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. Review of Educational Research,68, 350-386.
12. Khusainova R, Shilova Z. and Curteva O. (2016). Selection of Appropriate Statistical Methods for Research Results Processing. Mathematics Education, 11(1),303-315.
13. McMillan, J. H., & Schumacher, S. (2010). Research in Education Evidence-Based Inquiry.7th Edition. New Jersey: Pearson Education Inc.
14. Skidmore S. and Thompson B (2010). Statistical Techniques Used in Published Articles: A Historical Review of Reviews. Educational and Psychological Measurement, 70 (5), 777-795